Uncertainty quantification for fat-tailed probability distributions in aircraft engine simulations

Article (Accepted Version)

Uncertainty Quantification for Fat-Tailed Probability Distributions in Aircraft Engine Simulations

R. Ahlfeld and F. Montomoli

Uncertainty Quantification Lab, Imperial College, London, SW7 2AZ, U.K.

E. Scalas

Department of Mathematics, School of Mathematical and Physical Sciences, University of Sussex, U.K.

S. Shahpar

CFD Methods Group, Rolls-Royce, Derby, DE24 8BJ, U.K.

Rare event simulation is vital for industrial design, as some events, so-called Black Swans, can have fatal consequences despite their low probability of occurrence. Finding low probability events far off the mean design is a challenging task for realistic engineering models, because they are characterised by high computational demands, many input variables and often insufficient statistical information to build parametric probability distributions. Therefore, an adaptive and arbitrary Polynomial Chaos method, called SAMBA, is suggested in this work. SAMBA creates custom polynomial basis functions and grids based on statistical moments to avoid incorrect statistical assumptions. The contribution of this work is that it is derived how rare event simulation can conveniently be integrated into adaptive sparse grid methods by calculating Polynomial Chaos Expansions based on the statistical moments of truncated fat-tailed distributions. Moreover, the use of Tempered Alpha Stable Distributions is suggested to avoid discontinuous tail cut-offs. SAMBA is compared to other statistical methods in two industrial aircraft engine simulations: a simulation of transient cycle temperature in a turbine cavity and hot gas ingestion in the inter-wheel region. In both cases,

SAMBA agrees with previous results, but obtains them with lower computational effort.

μ_k	The k^{th} moment of a sample
M	Hankel matrix of moments
N	Number of available random samples
ξ	Input random variable
$w(\xi)$	Probability Density Function
Ω	Stochastic integration space
ψ_j	Univariate optimal orthogonal polynomial
Ψ_j	Multivariate linear combination of univariate optimal orthogonal polynomials
R	Cholesky decomposition matrix
r_{ij}	Entries of the Cholesky decomposition matrix
s_{ij}	Entries of the inverse of R
a_i,b_i	Three term recurrence coefficients
ω	Optimal Gaussian integration weights
J	Jacobi matrix
v_1	First eigenvector
U^i	Univariate Gaussian quadrature rule
N_U	Number of input distributions
m_{i_j}	Maximum adaptive order for a univariate quadrature rule
$I_j^{(k)}$	Smolyak index matrix
i_j	Entry of the Smolyak index matrix
$\|i\|$	Sum of a row of the index matrix
l	Level of Smolyak integration
A_S	Multivariate Smolyak quadrature
$\bar{\theta}$	Vector of sparse quadrature weights
η_i	Row i of the array of sparse collocation points

N_{sp} Number of sparse quadrature points

ξ^{i_j} Quadrature points used by U^{i_j}

α_k Fourier coefficients of Polynomial Chaos Expansion

$f(\xi)$ Model function

I. Introduction

Empirical data in many areas have fat-tailed or heavy-tailed probability distributions due to manufacturing or in service variations. Heavy-tailed distributions have been experimentally observed in fluid dynamics [1] and polymers [2]. They have been successfully utilised to model a variety of non-equilibrium dissipative systems, such as turbulent fluids [3, 4], anomalous diffusion [5], very stiff polymers [6], sub-recoil laser cooling [7], and the spectral random walk of a molecule embedded in a solid [8]. The most important area of application of heavy-tailed probability distributions is, however, with respect to Uncertainty Quantification (UQ) methods for the simulation of rare events.

For uncertainty propagation in engineering applications the normal distribution is most commonly used as input PDF to describe the uncertainty [9]. The reason is that major trends of many experimental and natural random errors can be described well with bell-shaped curves. Moreover, the approach is mathematically supported by the Central Limit Theorem, which states that the average of independent random variables from independent distributions has an approximately normal distribution. The possibility to propagate Gaussian distributions using efficient cubature formulae as described in [10] also adds to their popularity. However, it was first noted by Mandelbrot [11], and more recently by Taleb [12], that the assumption of normality can give a false sense of security, because it removes the possibility to account for rare events from the model. Events that are further away than five or more standard deviations from the mean are extremely unlikely for the normal distributions - more unlikely than they often are in reality. If such rare events have catastrophic consequences, they are called a Black Swan, as suggested by Taleb [13]. To obtain an accurate representation of the occurrence of rare events and Black Swans, heavy-tailed probability distributions instead of Gaussians need to be used [14].

Up to now, heavy-tailed distributions for rare events quantification have been mostly studied in mathematical finance to explain why market crashes occur more frequently than predicted by Gaussian models. They have been transferred to engineering applications by Montomoli [15, 16] only a few years ago. However, the authors used Monte Carlo simulations in combination with a meta-model to predict the impact of heavy-tailed distributions. The propagation of uncertainty caused by heavy-tailed probability distributions using a Non-Intrusive Polynomial Chaos (PC) method is still a neglected area of research in comparison to other aspects of UQ.

Non-intrusive polynomial chaos methods can be applied to propagate any probability distribution or random data set, as long as they have finite moments and their moment matrix is determinate in the Hamburger sense [17]. To circumvent the fact that heavy-tailed distributions have infinite moments, the distributions need to be truncated [18]. Smooth truncation was originally suggested by Koponen [19] to truncate Lévy flights. The truncated distributions have finite moments of all orders, while presenting fat tails up to a certain scale followed by a cut-off. This was done, for example, by Safta to develop basis functions for the fat tail of a lognormal distribution [20].

Here, the optimal Polynomial Chaos Expansion (PCE) for the truncated heavy-tailed distributions will be calculated using statistical moments of the input distributions. Moments are a quantitative measure in statistics with which any set of random samples (histogram), probability density function (PDF), or Cumulative Density Function (CDF) can be described uniformly [21, 22]. Most commonly, only the first four moments are used in engineering: mean, variance, skewness and kurtosis. Polynomial Chaos methods using the moments are the most general way to determine PCEs. This idea was first voiced and applied for stall flutter by Witteveen [23] and then adapted for geoscience applications by Oladyshkin [24–26]. Once the optimal PC basis has been determined using moments, various approaches like Galerkin projection, collocation or numerical integration [27] can be used to find the coefficients of the polynomial series describing the model output quantities.

In this work, a Sparse Approximation of Moment-Based Arbitrary Polynomial Chaos, SAMBA (PC), is used [17]. SAMBA combines a simple mathematical framework to determine the PCE using the statistical moments of the input distributions with an adaptive and anisotropic version of Smolyak's algorithm to alleviate the curse of dimensionality for multiple input variables. The

novelty of this work is that it is explained how PCEs can be developed in a suitable way for multiple smoothly truncated fat-tailed probability distributions so that reliable rare event simulations can be incorporated more easily into aircraft engine simulations.

In Section II, the basics of heavy-tailed probability distributions and their truncation for aircraft engines is explained. In Section III, it is described how SAMBA can be used to obtain a polynomial basis under an arbitrary probability measure and in particular for a truncated heavy-tailed distribution. In Section IV, the suggested methodology is demonstrated for a discontinuous cut-off in part IV A and for a smooth exponential cut-off in part IV B. In Section V, the method is first applied to propagate a heavy-tailed probability distribution through a CFD model for the transient cycle temperature of an aircraft engine in part V A. A validation of the used simulations for the engine transient was already presented by Montomoli [16] using a Monte Carlo Method with a meta-model. Second, the failure probability of a gas turbine due to hot gas ingestion is estimated based on a CFD simulation of the inter-wheel region of a gas turbine using a Meta-Model Monte Carlo, Importance Sampling, a full tensor Polynomial Chaos and SAMBA in part V B. The results demonstrate that the use of a moment-based method like SAMBA is particularly useful to perform rare event simulation for very large engineering models, because rare events can be characterised on top of standard UQ without having to change the model or UQ method.

II. Truncated Heavy-Tailed Probability Distributions

While there is no universal agreement about the terminology, the term heavy-tailed is prevalently considered as more general than fat-tailed and is used to refer to probability distributions which do not have exponentially bounded tails and therefore infinite moments [28]. Visually, heavy-tailed probability distributions can be recognised because they have higher peaks and kurtosis than for example the Gaussian distribution, see Figure 1. A large family of heavy-tailed distributions is given by Alpha stable Lévy distributions with shape factor $0 \leq \nu \leq 2$. Other important distributions are the Student-t distribution and the Cauchy distribution. The most distinguishing feature of all heavy or fat-tailed distributions is that their tails decay as $f(x) = x^k$ and not as $f(x) = k^x$, for any variable x and constant k. While this difference may seem negligible at a first glance of their

probability distribution in Figure 1, it has a strong effect on their moments. The raw moments of

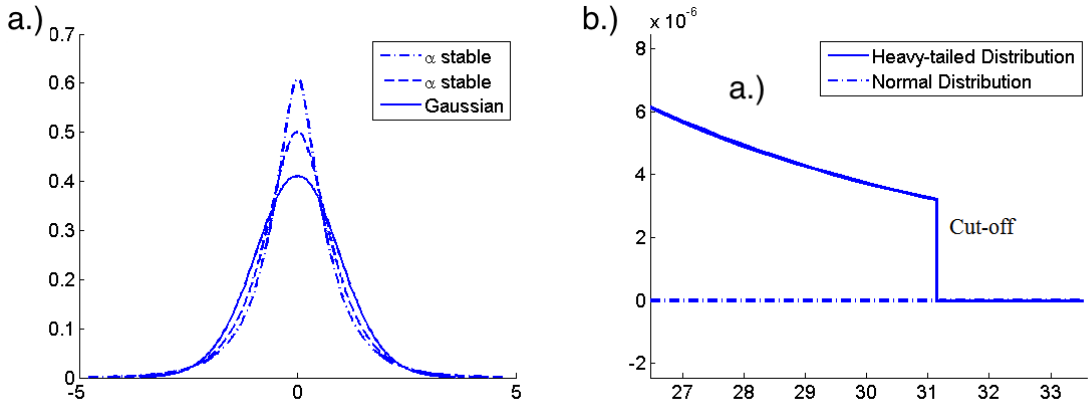

Fig. 1 a.) Alpha-stable distributions b.) Truncated tail of heavy-tailed distribution

a random variable ξ with probability density function $f(\xi)$ are defined as

$$\mu_k = \int_{\xi \in \Omega} \xi^k f(\xi) d\xi \tag{1}$$

If $f(\xi)$ decays as $1/x^k$ and not exponentially, the moments μ_k can become infinite. The application of polynomial chaos methods to propagate uncertainty caused by heavy-tailed distributions has so far been kept back in engineering by the fact that heavy-tailed distributions have infinite moments. This can however be circumvented by acknowledging that a cut-off can be imposed at the boundaries of physical systems or at the boundaries of their feasible operating domain. If this cut-off l_{cut} is also imposed on the distribution of the heavy-tailed random variable, it ensures the finiteness of the moments. However, a cut-off extremely far away from the mean can create a huge magnitude difference between the moments. This may cause the numerical method described in Section I to become ill-conditioned. The imposed cut-off can either be abrupt and discontinuous, as suggested by Mantegna and Stanley for alpha stable distributions [18] or smooth and continuous as described by Koponen [19] or Dubrulle [29]. While a continuous cut-off is more satisfactory from a mathematical point of view, because the intergral of the PDF remains exactly one, it does not make more sense from a physical perspective [29]. For example, in Lattice Boltzmann simulations the jumps of the particles are inevitably limited by the size of the box [4]. For this reason, a continuous but abrupt cut-off was suggested by Dubrulle [29]. However, Dubrulle remarks that this does not qualitatively affect the results. Since Dubrulle's results are merely valid for truncated Lévy distributions and

not heavy-tailed distributions in general, an abrupt truncation related to the physically realisable values of a random parameter is more advisable.

For the special case that the physical cut-off is so far off the mean that the moments become too high, it is possible to fit a tempered Lévy stable distribution to the original distribution and thereby impose a smooth exponential cut-off. The truncated distribution imitates the statistical properties of the non-truncated alpha stable Lévy distribution for smaller (finite) sample numbers and the sums of tempered stable random variables converge to normal random variables for a very huge (infinite) number of samples. Therefore, the usual Central Limit Theorem is not violated. For random variables z_i with identical distribution $f_L(z)$ the truncated Lévy distribution $L_T(x)$ is defined as the limit of the sum of the random variables z_i in the Poissonian stochastic process (as explained in Feller [30]), whose characteristic function $\varphi(k)$ obeys:

$$\ln\left(\varphi(k)\right) = \int_{-\infty}^{\infty} \left(e^{-ikz} - 1\right) f\left(z\right) dz \tag{2}$$

where f_L is defined as

$$f_L\left(z\right) = \begin{cases} C|z|^{-1-\nu} & z < 0 \\ Cz^{-1-\nu} & z > 0 \end{cases} \tag{3}$$

Using Lévy 's Inversion Formula the probability distribution $L_T(x)$ corresponding to the characteristic function can be calculated with

$$L_T\left(x\right) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \varphi(k) e^{-ikx} dx \tag{4}$$

III. Theoretical Basics of SAMBA PC

SAMBA consists of two components: the calculation of an individual Gaussian quadrature rule for each input distribution using statistical moments, and the sparsification to a anisotropic Smolyak quadrature rule to reduce the computational effort.

A. Optimal Gaussian Quadrature Rules for Arbitrary Random Variables

Statistical moments, like for example mean and variance, are a quantitative measure to describe the shape of any random variable. However, they are not widely used in applied uncertainty quantification. If PCE are based directly on moments, they can consider arbitrary random inputs in every

mathematical form without change in methodology: continuous, discrete PDFs, CDFs or random data sets/histograms [31]. The k^{th} raw moment μ_k of a continuous PDF $w(\xi)$ with $\xi \in \Omega$, or a set of N random samples $\zeta_1, ..., \zeta_N$, can be calculated with

$$\mu_k = \int_{\xi \in \Omega} \xi^k w(\xi) d\xi, \qquad \mu_k = \frac{1}{N} \sum_{i=1}^{N} \zeta_i^k. \tag{5}$$

The method suggested in this work to obtain the optimal Gaussian quadrature rules is fully described in Golub and Welsch and was derived by Mysovskih [32, 33]. Its advantage is that the optimal Gaussian quadrature points and weights can be computed directly from the Hankel matrix of moments. Here, only the the main relations are summarised. Full details are described in Ahlfeld [17]. The Hankel matrix of the moments is formed by

$$M = \begin{bmatrix} \mu_0 & \mu_1 & \cdots & \mu_p \\ \mu_1 & \mu_2 & & \mu_{p+1} \\ \vdots & & \ddots & \\ \mu_p & \mu_{p+1} & & \mu_{2p} \end{bmatrix}. \tag{6}$$

Since the Hankel matrix is positive definite, its Cholesky decomposition of $M = R^T R$ can be calculated. Using Mysovskih Theorem the optimal orthogonal polynomials ψ_j can be found as the entries of the inverse matrix R^{-1}. To avoid the costly inversion of the matrix M for larger systems, Rutishauser's formulas [34] can be used to obtain the polynomial coefficients of the orthogonal polynomials s_{ij} from the Cholesky matrix entries r_{ij}. With these relationships the three-term recurrence coefficients a_j and b_j in terms of r_{ij} can be calculated as

$$a_j = \frac{r_{j,j+1}}{r_{j,j}} - \frac{r_{j-1,j}}{r_{j-1,j-1}} \qquad b_j = \frac{r_{j+1,j+1}}{r_{j,j}}. \tag{7}$$

The recurrence coefficients a_j and b_j can be rearranged into a symmetric positive definite tridiagonal Jacobi matrix J, whose eigenvalues are the optimal Gaussian collocation points. The optimal Gaussian weights can be found as the first component of the normalized eigenvector v_1 corresponding to the first eigenvalue of J [33].

B. Sparsifying Moment-Based Gaussian Quadrature Rules

For multiple input distributions, the found Gaussian quadrature rules have to be combined using a tensor product of the individual optimal points and weights. Thus, the computational cost grows

exponentially with the number of inputs [35]. For more than 5 input distributions, the use of PC without sparsification becomes infeasible for industrial simulations with long run times. SAMBA, therefore, uses a sparsified quadrature rule based on Smolyak's Algorithm [36] to reduce the computational cost. Most commonly Smolyak's formulas are used based on polynomial interpolation at the extrema of the Chebyshev polynomials (Clenshaw-Curtis nodes) [35]. To obtain an anisotropic and adaptive quadrature rule, a sparse rule for an arbitrary combination of different uni-variate Gaussian quadrature rules optimal for the respective individual input distributions has to be formulated. For N_u sequences of one-dimensional quadrature rules $\{U^{i_j}\}_{j=1,...,N_u}$

$$U^{i_j} = \sum_{k=1}^{m_{i_j}} f(\xi_k^{i_j})\omega_k^{i_j}, \tag{8}$$

where m_{i_j} $j \in \{1, ..., N_u\}$ is the maximum adaptive order for each univariate quadrature rule, the multivariate sparse Smolyak quadrature rule $A_S\left(N_u + l, N_u\right)$ is

$$A_S = \sum_{\substack{l+1 \\ \leq |i| \leq \\ l+N_u}} (l-1)^{l+N_u-|i|} \begin{pmatrix} N_u - 1 \\ \\ l + N_u - |i| \end{pmatrix} \otimes_{k=1}^{N_U} U^{i_k} \tag{9}$$

where l is the level, which controls the accuracy of the result. For the same number of uncertain inputs N_u an increased level l improves the accuracy, but increases the computational effort [37]. In general, only low levels (first and second) result in a significant reduction of the computational effort. The term $|i|$ is the norm of the vector $i = \{i_1, ..., i_{N_u}\}$ which stands for the sum of a row j of the index matrix $I_j^{(k)}$. To differentiate the sparse grid points from the uni-variate optimal points, the i^{th} row of the sparse collocation points array is referred to as η_i. The Fourier coefficients of the resulting PCE α_k can be numerically approximated with

$$\alpha_k = \frac{\sum_{i=1}^{N_{Sp}} f\left(\eta_i\right) \Psi_k\left(\eta_i\right) \theta_i}{\sum_{i=1}^{N_{Sp}} \Psi_k\left(\eta_i\right) \theta_i} \tag{10}$$

where N_{Sp} is the number of sparse points. The normalisation of the polynomials at this stage is not required, if orthonormal polynomials were determined before. The PC Fourier coefficients are required to sample the PCE for the occurrence and probability of rare events. If only the moments of the posterior distribution $E\left[f^k\right]$ are required, they can be obtained faster by using the sparse

quadrature formulas directly on the formula for the k^{th} moment:

$$E\left[f_k\right] = \sum_{i=1}^{N_{Sp}} \left(f\left(\eta_i\right) - E\left[f^{k-1}\right]\right)^k \theta_i. \tag{11}$$

where f is a non-linear model and the sparse collocation points and weights are η_i and θ_i.

IV. Two Validation Examples for PCEs of Fat-Tailed PDFs

In order to demonstrate the proposed polynomial method it is applied to a simple one dimensional case, for which the model is given by the analytical formula

$$g\left(\tilde{t}\right) = 0.05\left(\tilde{t} - 10\right)^3 + 3\left(\tilde{t} - 10\right) + 120 \tag{12}$$

where $\tilde{t}$ consists of a deterministic and stochastic part: $\tilde{t} = t + \xi$, $t \in R$ and ξ is a random variable with three degrees of freedom Student-t probability distribution ($\nu = 3$). Eq. (12) is chosen as a simplification of the transient gas turbine temperature profile used in the real engineering example in Section V. Figure 2 shows the optimal Gaussian and fat-tailed collocation points for the halved distribution that occur because the time is assumed to be non-negative.

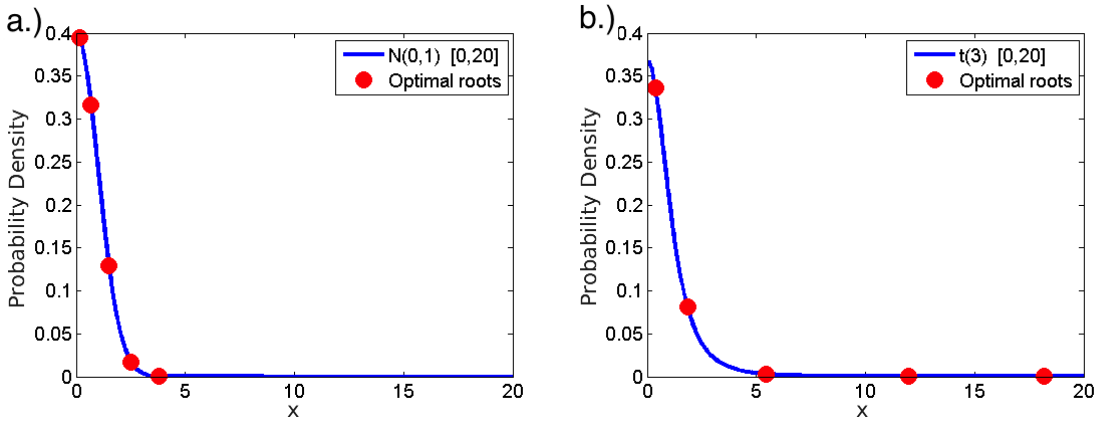

Fig. 2 Optimal collocation points for a.) Gaussian Distribution and b.) Student-t Distribution

A. Numerical Validation with Discontinuously Truncated T-Distribution

For a heavy-tailed t-distribution with only 3 degrees of freedom, only the mean and variance are finite. Skewness, kurtosis and all higher moments are undefined. Since the finiteness of the moments is a necessary condition to apply arbitrary polynomial chaos, the distributions has to be truncated.

In this case, it will be assumed that the stochastic space of ξ is limited between 0 and 20. The truncated heavy-tailed distributions have finite moments, so that SAMBA can be applied. Particularly noteworthy is that SAMBA can even propagate the half probability distribution created at $t = 0$ and that the optimal collocation points for this case are entirely positive, as shown in Figure 3. The red arrows in both graphs have the same lengths at the same position in time. The results are

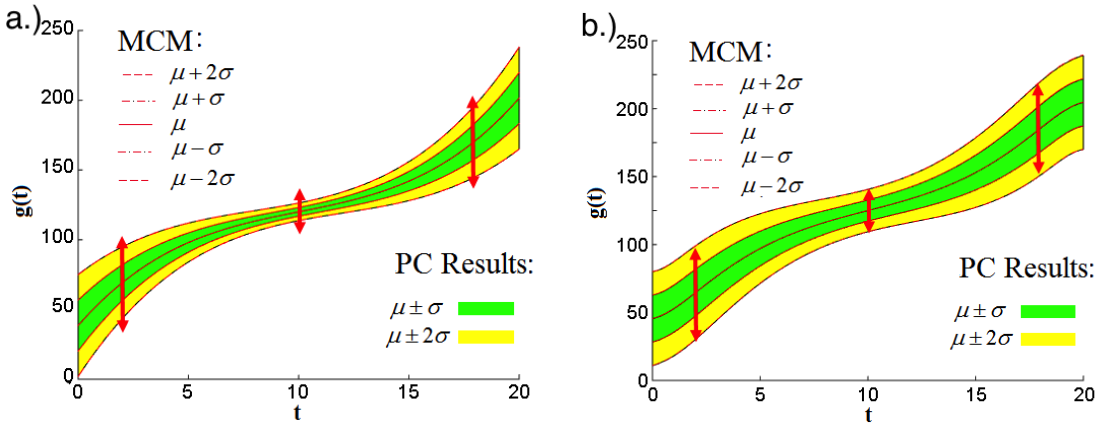

Fig. 3 Output variation for a.) Gaussian and b.) heavy-tailed input distribution

validated with a Monte Carlo Method with 10^6 samples. The samples for the MCM are drawn from a heavy-tailed t-distribution with three degrees of freedom, $\nu = 3$, which has also been truncated to the interval $[0, 20]$.

B. Obtaining the PCE for a Tempered Alpha Stable Distributions

In case a smooth exponential cut-off is needed, the characteristic function $\varphi_X(t) = e^{\phi(X)}$ can be developed analytically by integrating an exponential cut-off into the probability distribution f_L from Eq. (3):

$$L_T(z) = L(z)\, e^{-\lambda z} = \begin{cases} C|z|^{-1-\nu} \cdot pe^{-\lambda|z|} & z < 0 \\ Cz^{-1-\nu} \cdot qe^{-\lambda z} & z > 0 \end{cases} \tag{13}$$

where λ is the variable used to control the cut-off. The characteristic function $\varphi_X(t) = e^{\phi(X)}$ can be derived as described by Nakao [38]

$$\phi_X(t) = C\Gamma(-\nu)\left[\left(\lambda^2 + t^2\right)^{\nu/2} \cos(\nu\theta) + i(q-p)\sin(\nu\theta) - \lambda^\nu\right] \tag{14}$$

for $0 < \nu < 1$ and with $\theta = \arctan\left(\frac{t}{\lambda}\right)$ and

$$\phi_X(t) = C\Gamma(-\nu)\left[q(\lambda + it)^\nu + p(\lambda - it)^\nu - \lambda^\nu - i\nu\lambda^{\nu-1}(q-p)\,t\right] \tag{15}$$

for $1 < \nu < 2$.

By using the expansion: $q(\lambda + it)^\nu + p(\lambda - it)^\nu = \left(\lambda^2 + t^2\right)^{\nu/2}\cos(\nu\theta) + i\beta\sin(\nu\theta)$ (where $\beta = q - p$ describes the symmetry) it can be seen that in the limit $\lambda \to 0$, the tempered stable distribution converges to the stable distribution. Unfortunately, an analytical expression for the probability distribution is not attainable for more than a few exceptional cases [38]. The probability density $L_T(x)$ from Eq. (4) can, however, be obtained numerically. By applying adaptive numerical integration suitable for oscillating integrals (like, for example, Gauss-Kronrod Quadrature) to Eq. (14) or (15) accurate numerical values for $f_X(x)$ can be calculated [19]. This was done in Figure 4 for various values of the cut-off parameter λ from purely fat-tailed (Lévy distribution) to fully Gaussian tails.

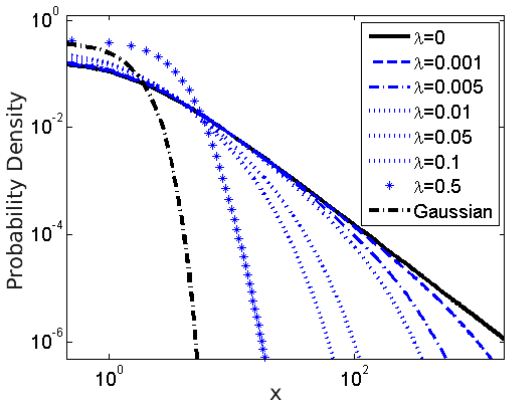

Fig. 4 Tails of tempered alpha stable distributions for various cut-off parameters.

The parameters of the characteristic function of a tempered Alpha Stable Lévy distribution as characterised by Eq. (14) or (15) can be determined by using a Maximum-Likelihood approach. It is important to pre-set the cut-off parameter λ to a desired value before performing the distribution fitting, because its value influences the overall behaviour of the distribution. A fitted distribution is shown in Figure 5. If the physical constraints on the stochastic parameter are known beforehand λ should be chosen such that the tails of the tempered distribution follow the tails of the not-truncated Lévy distribution in the relevant physical domain.

For example: by choosing the cut-off parameter as $\lambda = 0.05$ and fitting a tempered alpha

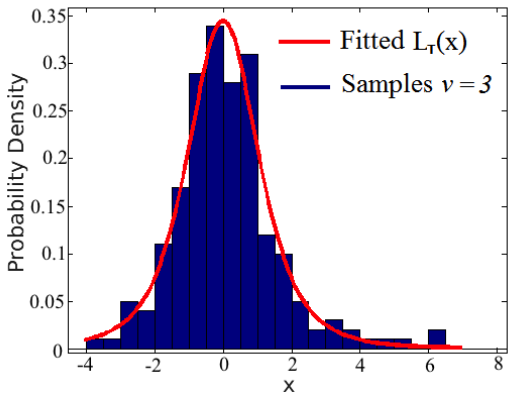

Fig. 5 Alpha stable distribution fitted to fat-tailed Student-t samples

stable distribution to the histogram of 10000 samples obtained from a student-t distribution with 3 degrees of freedom as shown in Figure 4, the following parameters can be obtained: $C = 0.411569$, $\nu = 1.5268$ and $\beta = 0$, thus $p = 1$ $q = 1$. The optimal orthogonal polynomials for this exponentially tempered heavy-tailed distribution are approximately:

$$\begin{aligned} &1 \\ &x \\ &x^2 - 3.1525 \\ &\mathrm{x}^3 - 262.52x \\ &x^4 - 2393.91x^2 + 6719.26 \\ &\mathrm{x}^5 - 4898.38\mathrm{x}^3 + 664205\mathrm{x} \\ &\mathrm{x}^6 - 771.769\mathrm{x}^4 + 7413360x^2 - 18943630 \end{aligned} \tag{16}$$

V. Two Engineering Examples

In this section SAMBA will be applied to model rare events in two gas turbine simulations: Stochastic variation of metal temperature during a transient and hot gas ingestion.

A. Example I: Stochastic Variation of Metal Temperature during a Transient

The details of the subsequent engineering model are explained in full in Montomoli [16] and are briefly summarised in the following. Running gas turbines follow prescribed missions for which the occurring metal temperatures are predicted using deterministic computational models. Figure

6 shows a representative flight cycle for the time-dependent metal temperature in a region of the turbine stator well, as it occurs in the transition process from a lower shaft rotation speed ($U = 0.226U_M$) to a higher speed ($U = U_M$). The transition phase is marked by two acceleration ramps: A slower (first) and faster (second) ramp. The slower ramp takes about 270s and the faster 9s.

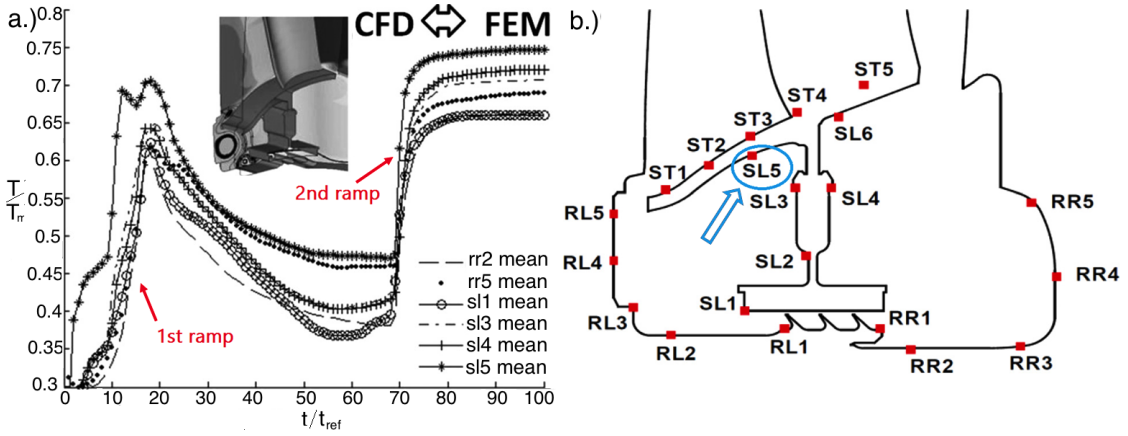

Fig. 6 a.) Temperatures curves at different rotor cavity locations b.) Temperature measurement points

The temperature curves in the right graph of Figure 6 were obtained from a coupled CFD/FEM simulation for the solid 3D model. It is based on a real aircraft engine and the boundary conditions were assigned matching the engine's instantaneous operating conditions. Several locations are used for monitoring the temperature as described by Amirante [39] and as shown in the right graph of Figure 6. The temperature has been non-dimensionalised using the maximum value of annulus temperature T_M pertaining to the high speed following the cycle and the time is non-dimensionalised with the overall mission time of the transition $t_{ref} = 4000s$. Under operating conditions, the mission cycle is exposed to stochastic variations. In the test bed, this variation was shown to be in the order of several seconds. In Montomoli [16] it was approximated with a Gaussian distribution with standard deviation $\sigma = 1.125t_{ref}\%$. Under real flight conditions the uncertainty is expected to be even higher. The importance of considering time variation was shown in several technical reports, see for example Section 3.7 in Agard [40] or [41]. For example, a fast shut-down can lead to an engine life reduction of 200 hours [16]. In a similar way, high transient temperature gradients can also reduce the engine life and a metal variation of 20K changes the life about 50%.

To test the sensitivity of the model towards fatter tails SAMBA is used here. For this case benchmark results are available from Montomoli [16], who showed that a small variation in the tail of the probability distribution describing mission time uncertainty, namely from Gaussian to Student-t with three degrees of freedom, is most strongly reflected where the temperature gradients are highest. The effect of fat tails was simulated by performing a Monte Carlo Method (MCM) with a Student-t temperature distribution in sensor $SL5$. Because a direct standard MCM was infeasible due to the complexity of the CFD model, Montomoli calculated a response surface and sampled this surface 10^8 times instead of running the entire model that often. To avoid non-physical

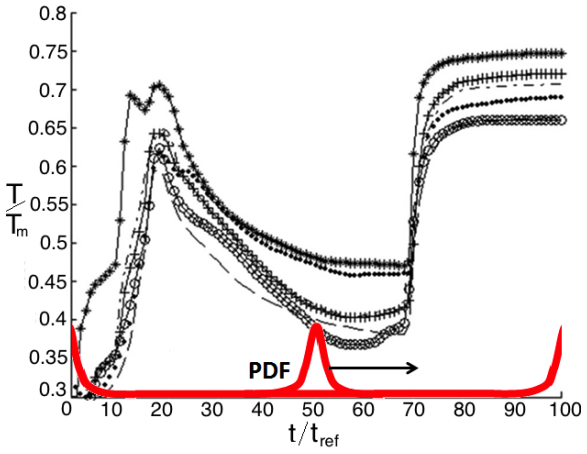

Fig. 7 PDF movement on truncated time domain

time variations, negative time or infinitely large time, the time variations are simulated by making random draws from a time array reaching from 0 to 100 t/t_{ref} using a Student-t distribution with three degrees of freedom, such that the current time is always the mean, but the tails of the used PDF are cut-off if they reach either $t = 0$ or $t = 100$, as depicted in Figure 7. Exactly the same domains are used to calculate the orthogonal polynomials for SAMBA. While a meta-model approach is computationally much less expensive than a direct MCM, it comes with certain disadvantages. First, sampling the response surface requires just as many runs as a brute-force MCM: 10^8 runs, plus the recreation of the response surface required 23 CFD simulations in this case. Second, the creation of a trustworthy response surface usually requires several checks so that the overall computational effort is higher than for SAMBA. SAMBA is therefore more efficient and needs significantly fewer model runs to evaluate the moments, especially if considering that only 6 CFD simulations and 6

model evaluation are required in total for a fifth order approach for only one uncertain variable. Figure 8 shows that SAMBA provides exactly the same results with lower computational cost.

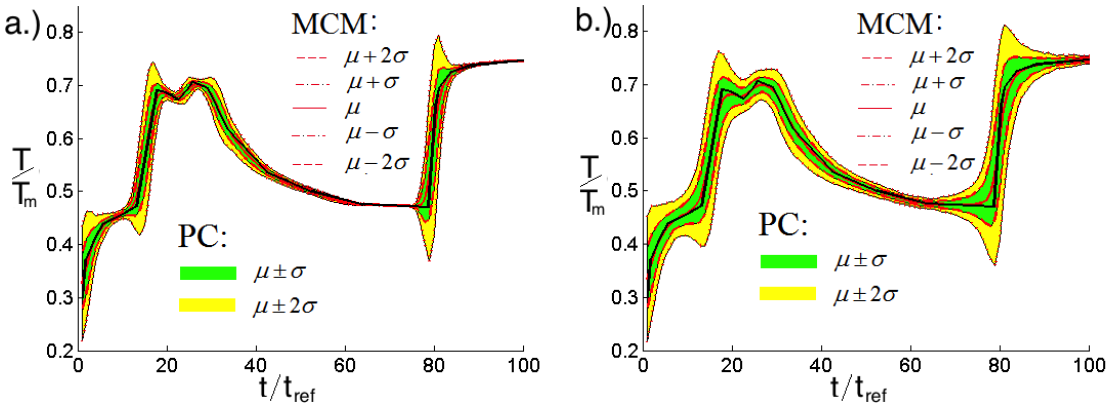

Fig. 8 SAMBA and MCM comparison for a.) Gaussian and b.) Student-t input distribution

Figure 8 a.) shows the results obtained for an input Gaussian distribution, and Figure 8 b.) for a Student-t distribution. The iso-contours of the mean value of the transient temperature with +/-1 and 2 standard deviations are shown against the dashed lines obtained by a Monte Carlo Simulation with a response surface.

B. Example II: Hot Gas Ingestion in Gas Turbines

One of the most important problems faced in modern gas turbine design is the ingestion of hot mainstream gas into the inter-wheel spaces between the turbine rotors and the spacer [42]. The reason is that the ingress of hot gas into the lower cavity can result in dangerous damage to the turbine rim, blade roots and discs with a reduction of component life [15]. To prevent hot gas ingestion, a stable 'cold' purge flow is bled from the compressor and used to cool the cavity. However, being removed from the compressor, this colder air has been compressed (using mechanical energy), but has a minimum contribution to the turbine power output because it has not been energized by the combustion energy. This limits the amount of coolant that can be used because it directly affects the engine's efficiency, which leads to an increase in fuel consumption and CO_2 emissions. Therefore, a challenging task is to balance the purge flow across the spacer. Moreover, the presence of random gaps between the stator and the diaphragm complicates the problem with unwanted secondary air flow system leakages. In this work, state-of-the-art methods

have been used for the CFD simulations in combination with SAMBA to calculate the statistics of the temperature in the lower cavity as well as the probability of hot gas ingestion. The most important aleatory factors influencing the probability are the sizes of gap 1 and gap 2 as displayed in Figure 9. How a stochastic variation of ±50% of the gap sizes influences the probability of failure

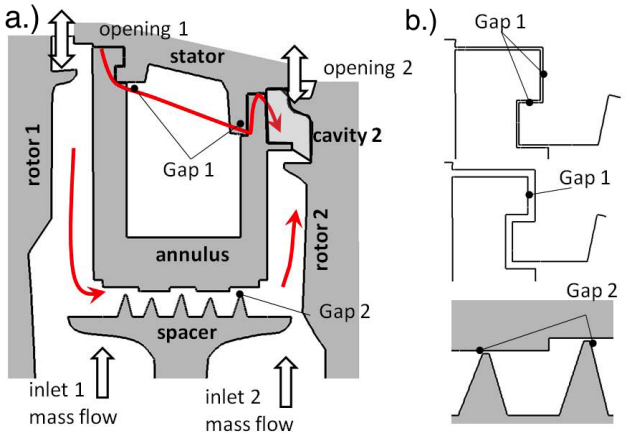

Fig. 9 Illustration of a.) hot gas ingestion mechanism b.) two gaps varied in simulation

was investigated by Montomoli [15] using a Meta-Model Monte Carlo Simulation based on 25 full CFD runs. The CFD simulation used in the following is the same as performed by Montomoli in the previous work: a quasi-3D simulation of the cavity region with the CFD solver CFX and the k-ω SST turbulence closure [43]. The mesh used contains 0.502M elements, which was shown to be sufficient, and it is comparable to the industrial standard in similar cases. All temperature values are normalised by the maximum temperature in the free stream gas path of a representative 32MW machine. Further details regarding the boundary conditions and the setting of reference temperatures can be found in Montomoli [15] and are based on realistic gas turbine values and measured distributions in real gas turbines. The simulation and results presented here, however, are only an example of the application of SAMBA and the comments and conclusions do not reflect any specific gas turbine. Under nominal operating conditions the cavity is correctly purged and the normalised temperature is 0.86, which is well below the critical value of $T_{crit} = 0.92$. However, the gap sizes can vary from their mean value due to manufacturing variation. The standard deviation of this uncertainty will be assumed to be $\sigma = 6\%$ in the following. For the application of SAMBA two histograms based on 1000 samples are drawn from Gaussian distributions for the two gaps

sizes as displayed in the schematic overview in Figure 10. We assume that the true underlying distributions are Gaussian, although only a limited amount of samples are available. The use of a

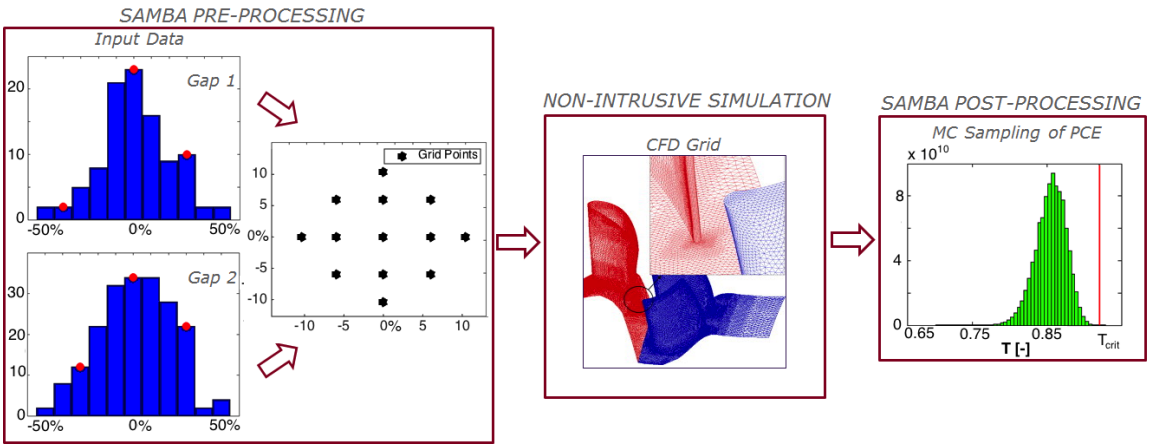

Fig. 10 Schematic overview of SAMBA application for hot gas ingestion

smaller set of measurement data, is not only more realistic, but also demonstrates that SAMBA can propagate real measurement data without needing PDF fitting. The optimal collocation points for each individual input histogram are computed and combined into a nearly Gaussian looking sparse Smolyak mesh. Moreover, also two Student-t distributions are fitted to the data to perform rare event analysis with fat-tailed distributions. The grids in Figure 11 show that the Gaussian grid takes only the main bulk of the distributions into account, whereas the fat-tailed grids dedicate one point to explore the failure regions. The Gaussian mesh is not only shown in the schematic, but also

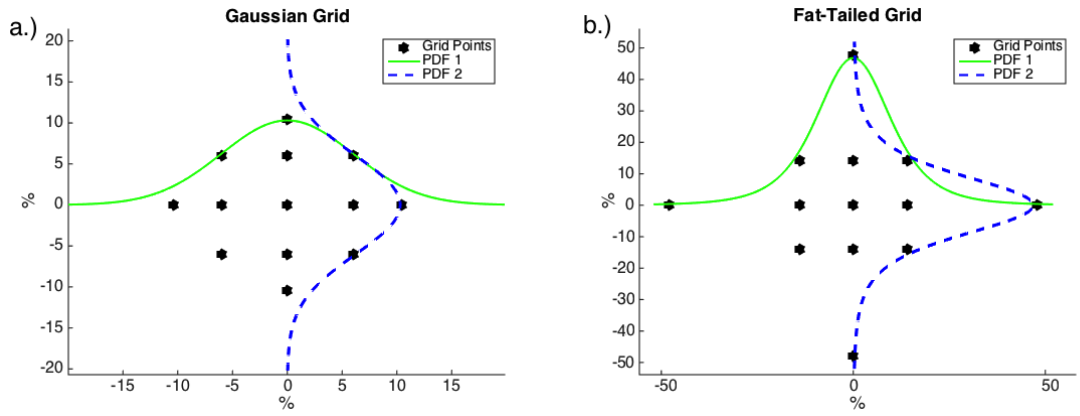

Fig. 11 SAMBA grids for a.) Gaussian and b.) KoBoL input distributions

in Figure 12 a.) in front of the contour lines of the response surface created through varying the

two gap sizes. The response surface in Figure 12 b.) was set-up for the Monte Carlo Meta-Model, which is also displayed in the figure. Most importantly, it can be seen that the manually set-up grid for the meta-model overestimates the relevant region to evaluate the effect of the inputs, while it under-samples the important region. Choosing the relevant region and an appropriate sampling rate is the main difficulty in manually designing meta-models. SAMBA on the other hand provides an optimal Design of Experiment (DoE) based on the moments. Practically, the resulting Smolyak grid represents a list of various gap sizes for which the CFD code needs to be executed. From the 13 CFD code results, the moments of the stochastic behaviour of the temperature can be derived. Moreover, by sampling the PCE the shape of the posterior PDF can be obtained. Since the lowest

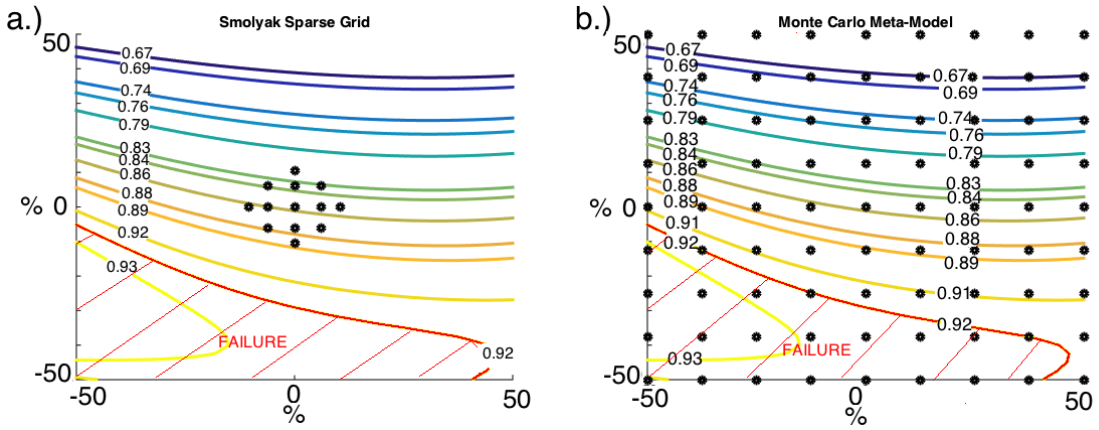

Fig. 12 Contour plots of a.) SAMBA grid and b.) Meta-Model grid

level Smolyak model only contains 5 points close to the mean, the second level Smolyak grid was used. To calculate the probability of failure, the PCE based on a full tensor with 25 simulations, a Monte Carlo with 25 simulations, direct importance sampling, and a Smolyak sparse grid with 13 simulations were compared. To obtain estimates of the probability distributions the meta-model and the PCEs were sampled with 10^8 samples. For importance sampling, the sampling distributions for the gap sizes were changed from $N(0, 6)$ to $N(37, 6)$ for gap 2 and $N(17, 6)$ for gap 1. This creates samples mainly in the estimated region of failure indicated by the red stripes in Figure 12. In this way, the probability of failure P_f of the response function of the CFD model $f(x, y)$ can be

estimated orders of magnitudes more efficiently by calculating

$$P_{f,IS} = \frac{1}{N}\sum_{i=1}^{N} f\left(x_i, y_i\right)\frac{w\left(x_i, y_i\right)}{g\left(x_i, y_i\right)} = \frac{1}{N}\sum_{i=1}^{N} f\left(x_i, y_i\right) b\left(x_i, y_i\right) \tag{17}$$

where $w\left(x, y\right)$ and $g\left(x, y\right)$ are distributed as

$$w\left(x, y\right) \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 6 & 0 \\ 0 & 6 \end{bmatrix}\right) \qquad g\left(x, y\right) \sim N\left(\begin{bmatrix} 37 \\ 17 \end{bmatrix}, \begin{bmatrix} 6 & 0 \\ 0 & 6 \end{bmatrix}\right). \tag{18}$$

While Importance Sampling is of course less efficient for a mere two uncertain input variables, it becomes quickly more attractive as the dimension of the problem increases, because it is not affected by the curse of dimensionality. The results of all methods are summarised in Table 1. The effect

Table 1 Probability of critical hot gas ingestion

Method	Order	Runs	Probability of Failure $N(\xi)$	Probability of Failure $t(\xi)$
Importance Sampling		10^3	$2.2495 \cdot 10^{-4}$ %	3.57 %
Meta-Model MCS		25	$2.04557 \cdot 10^{-4}$ %	3.22 %
Tensor Quadrature	$p = 4$	25	$2.05017 \cdot 10^{-4}$ %	3.21 %
SAMBA PC	$l = 2$	13	$2.02901 \cdot 10^{-4}$ %	3.04 %

that the fat tails of the input distribution have on the output PDF is shown in Figure 13, where T_{crit} indicates the failure region. The results show that sparse grids can be used to predict rare

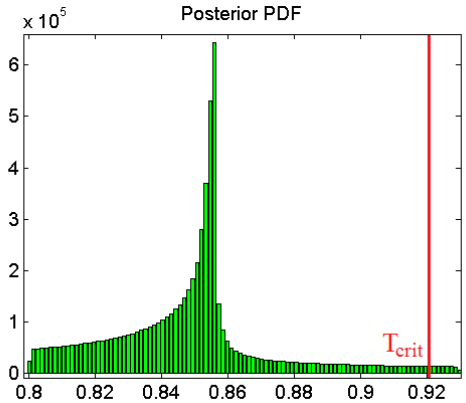

Fig. 13 Posterior PDF of engine temperate for fat-tailed inputs

events and probability of failures. The advantage of using SAMBA is that that the lower moments

of the posterior probability distribution can be obtained quite efficiently for larger problems where only small data sets and no PDFs are available.

VI. Conclusion

This work investigated the use of Polynomial Chaos sparse grids developed from the statistical moments of fat-tailed distributions for rare event simulation in aircraft engines models and compared it to efficient sampling approaches like meta-model Monte Carlo and Importance Sampling. To generate sparse Gaussian collocation points and weights, the method SAMBA was used, because it can simultaneously and without adaption integrate arbitrary small non-Gaussian data sets into the grid generation. SAMBA and the alternative methods were applied to two industrial engineering test cases: an aircraft engine temperature transient simulation and hot gas ingestion. For the chosen test cases, the rare event simulation could be performed more easily and efficiently using SAMBA. The reason being that sparse grids generated from the optimal Gaussian collocation points and weights of fat-tailed probability distributions can not only propagate the main shape of the distribution but also the fatter tails. Since it is a necessary requirement to truncate fat-tailed distributions in order to develop polynomial basis functions, a new approach based on fitting Alpha Stable Tempered Distributions was also suggested. This approach makes use of a smooth exponential cut-off to keep the moments finite, but the tails can be kept fat in an arbitrary long region. For problems with distinct physical limitations, the distributions can of course also be truncated discontinuously. Overall, it was demonstrated that Gaussian sparse grids obtained from fat-tailed distributions using SAMBA result in identical results as Meta-Model Monte Carlo and Importance Sampling for the probability of failure, but need fewer CFD simulations as the collocation points are placed more favourably to include the tail regions.

[1] Solomon, T. H., Weeks, E. R., and Swinney, H. L., "Observation of Anomalous Diffusion and Levy Flights in a Two-Dimensional Rotating Flow," *Physical Review Letters*, Vol. 71, No. 24, 1993, p. 3975.

[2] Ott, A., Bouchaud, J. P., Langevin, D., and Urbach, W., "Anomalous Diffusion in Living Polymers: A genuine Levy Flight?" *Physical Review Letters*, Vol. 65, No. 17, 1990, pp. 2201–2204, doi:10.1103/PhysRevLett.65.2201.

[3] Shlesinger, M. F., Klafter, J., and J. West, B., "Levy Walks with Applications to Turbulence and

Chaos," *Physica A: Statistical Mechanics and its Applications*, Vol. 140, No. 1-2, 1986, pp. 212–218, doi:10.1016/0378-4371(86)90224-4.

[4] Hayot, F. and Wagner, L., "Robustness of Vortex Streets," *Physical Review E*, Vol. 49, No. 1, 1994, pp. 470–473,
doi:10.1103/PhysRevE.49.470.

[5] Solomon, T., Weeks, E. R., and Swinney, H. L., "Chaotic Advection in a Two-Dimensional Flow: Levy Flights and Snomalous Diffusion," *Physica D: Nonlinear Phenomena*, Vol. 76, No. 1-3, 1994, pp. 70–84, doi:10.1016/0167-2789(94)90251-8.

[6] Moon, J. and Nakanishi, H., "Self-Avoiding Levy Walk: A Model for Very Stiff Polymers," *Physical Review Letters*, Vol. 42, No. 6, 1990, pp. 3221–3224,
doi:10.1103/PhysRevA.42.3221.

[7] Bardou, F., Bouchaud, J. P., Emile, O., Aspect, A., and Cohen-Tannoudji, C., "Subrecoil Laser Cooling and Levy Flights," *Physical Review Letters*, Vol. 72, No. 2, 1994, pp. 203–206,
doi:10.1103/PhysRevLett.72.203.

[8] Zumofen, G. and Klafter, J., "Spectral Random Walk of a Single Molecule," *Chemical Physics Letters*, Vol. 219, No. 3, 1994, pp. 303–309.

[9] Kmenta, J. and Rafailzadeh, B., *Elements of Econometrics*, University of Michigan Press, St. State, 1997.

[10] Lu, J. and Darmofal, D. L., "Higher-dimensional integration with Gaussian weight for applications in probabilistic design," *SIAM Journal on Scientific Computing*, Vol. 26, No. 2, 2004, pp. 613–624.

[11] Mandelbrot, B. B., *The Fractal Geometry of Nature*, Freeman, New York, USA, 1982.

[12] Suárez-Lledó, J., *The Black Swan: The Impact of the Highly Improbable*, Vol. 25, Random House Publisher, New York, USA, 2011,
doi:10.5465/AMP.2011.61020810.

[13] Taleb, N. N., *The Black Swan: The Impact of the Highly Improbable*, Random House and Penguin, New York, USA, 2007.

[14] Viswanathan, G. M., Fulco, U. L., Lyra, M. L., and Serva, M., "The Origin of Fat-Tailed Distributions in Financial Time Series," *Physica A: Statistical Mechanics and its Applications*, Vol. 329, No. 1, 2003, pp. 273–280.

[15] Montomoli, F. and Massini, M., "Gas Turbine and Uncertainty Quantification: Impact of PDF Tails on UQ Predictions, The Black Swan," in "Proceedings of ASME Conference GT2013," , 2013.

[16] Montomoli, F., Amirante, D., Hills, N., Shahpar, S., and Massini, M., "Uncertainty Quantification, Rare

Events, and Mission Optimization: Stochastic Variations of Metal Temperature During a Transient," *Journal of Engineering for Gas Turbines and Power*, Vol. 137, 2014, pp. 42–101, doi:10.1115/1.4028546.

[17] Ahlfeld, R., Belkouchi, B., and Montomoli, F., "SAMBA: Sparse Approximation of Moment-Based Arbitrary Polynomial Chaos," *Journal of Computational Physics*, Vol. 319, 2016, pp. 1–29, doi:http://dx.doi.org/10.1016/j.jcp.2016.05.014.

[18] Mantegna, R. N. and Stanley, H. E., "Stochastic Process with Ultrasound Convergence to a Gaussian Truncated Levy Flight," *Physical Review Letters*, Vol. 73, No. 22, 1994, p. 2946.

[19] Koponen, I., "Analytical Approach to the Problem of Convergence of Truncated Levy Flights Towards the Gaussian Stochastic Process," *Physical Review E*, Vol. 52, No. 1, 1995, pp. 1197–1199, doi:doi:10.1103/physreve.52.1197.

[20] Safta, C., Debusschere, B. J., Najm, H. N., and Sargsyan, K., "Advanced methods for uncertainty quantification in tail regions of climate model predictions." Tech. rep., Sandia National Laboratories, 2010.

[21] Ernst, O. G., Mugler, A., Starkloff, H.-J., and Ullmann, E., "On the Convergence of Generalized Polynomial Chaos Expansions," *Mathematical Modelling and Numerical Analysis*, Vol. 46, No. 02, 2012, pp. 317–339.

[22] Soize, C. and Ghanem, R., "Physical Systems with Random Uncertainties: Chaos Representations with Arbitrary Probability Measure," *SIAM Journal on Scientific Computing*, Vol. 26, No. 2, 2004, pp. 395–410, doi:10.1137/S1064827503424505.

[23] Witteveen, J. a. S., Sarkar, S., and Bijl, H., "Modeling Physical Uncertainties in Dynamic Stall Induced Fluid-Structure Interaction of Turbine Blades Using Arbitrary Polynomial Chaos," *Computers and Structures*, Vol. 85, No. 11-14, 2007, pp. 866–878, doi:10.1016/j.compstruc.2007.01.004.

[24] Oladyshkin, S., Class, H., Helmig, R., and Nowak, W., "A Concept for Data-Driven Uncertainty Quantification and its Application to Carbon Dioxide Storage in Geological Formations," *Advances in Water Resources*, Vol. 34, No. 11, 2011, pp. 1508–1518, doi:http://dx.doi.org/10.1016/j.advwatres.2011.08.005.

[25] Oladyshkin, S., Schröder, P., Class, H., and Nowak, W., "Chaos Expansion based Bootstrap Filter to Calibrate CO2 Injection Models," *Energy Procedia*, Vol. 40, 2013, pp. 398–407, doi:10.1016/j.egypro.2013.08.046.

[26] Oladyshkin, S., Class, H., Helmig, R., and Nowak, W., "An Integrative Approach to Robust Design and Probabilistic Risk Assessment for CO2 Storage in Geological Formations," *Computational Geosciences*, Vol. 15, No. 3, 2011, pp. 565–577, doi:10.1007/s10596-011-9224-8.

[27] Hosder, S., Walters, R. W., and Balch, M., "Point-Collocation Nonintrusive Polynomial Chaos Method for Stochastic Computational Fluid Dynamics," *AIAA Journal*, Vol. 48, No. 12, 2010, pp. 2721–2730, doi:10.2514/1.39389.

[28] Pisarenko, V. F., Rodkin, M. V., and Rodkin, M. V., *Heavy-Tailed Distributions in Disaster Analysis*, Springer Science & Business Media, Berlin, 2010.

[29] Dubrulle, B. and Laval, J.-P., "Truncated Levy Laws and 2D Turbulence," *The European Physical Journal B - Condensed Matter and Complex Systems*, Vol. 4, No. 2, 1998, pp. 143–146, doi:10.1007/s100510050362.

[30] Feller, W., *An Introduction to Probability Theory and Its Applications*, Russian Academy of Sciences, Branch of Mathematical Sciences, 1968.

[31] Oladyshkin, S. and Nowak, W., "Data-Driven Uncertainty Quantification Using the Arbitrary Polynomial Chaos Expansion," *Reliability Engineering and System Safety*, Vol. 106, 2012, pp. 179–190, doi:10.1016/j.ress.2012.05.002.

[32] Mysovskikh, I. P., "On the Construction of Cubature Formulas with Fewest Nodes," *Doklady Akademii Nauk SSSR*, Vol. 178, No. 6, 1968, pp. 1252–1254.

[33] Golub, G. H. and Welsch, J. H., "Calculation of Gauss Quadrature Rules," *Mathematics of Computation*, Vol. 23, No. 106, 1968, pp. 221–221, doi:10.1090/S0025-5718-69-99647-1.

[34] Rutishauser, H., "On a Modification of the QD-Algorithm with Graeffe-Type Convergence," in "Zeitschrift für angewandte Mathematik und Physik," Vol. 13, 1963, pp. 493–496.

[35] Eldred, M. S. and Burkardt, J., "Comparison of Non-Intrusive Polynomial Chaos and Stochastic Collocation Methods for Uncertainty Quantification," in "Proceedings of the 47th AIAA Aerospace Sciences Meeting," , 2009, pp. 1–20.

[36] Smolyak, S. A., "Quadrature and Interpolation Formulas for Tensor Products of Certain Classes of Functions," *Doklady Akademii Nauk SSSR*, Vol. 4, No. 240-243, 1963, p. 123.

[37] Judd, K. L., Maliar, L., Maliar, S., and Valero, R., "Smolyak Method for Solving Dynamic Economic Models: Lagrange Interpolation, Anisotropic Grid and Adaptive Domain," *Journal of Economic Dynamics and Control*, Vol. 44, 2014, pp. 92–123,

doi:10.1016/j.jedc.2014.03.003.

[38] Nakao, H., "Multi-Scaling Properties of Truncated Levy Flights," *Physics Letters, Section A: General, Atomic and Solid State Physics*, Vol. 266, 2000, pp. 282–289, doi:10.1016/S0375-9601(00)00059-1.

[39] Amirante, D., Hills, N., and Barnes, C., "A Moving Mesh Algorithm for Aero-Thermo-Mechanical Modeling in Turbomachinery," *Wiley International Journal for Numerical Methods in Fluids*, Vol. 70, No. 9, 2012, pp. 1–14.

[40] Dudgeon, E. H., "Guide to the Measurement of the Transient Performance of Aircraft Turbine Engines and Components (Report No. AR-320)," Tech. rep., Advisory Group for Aerospace Research and Development, Neuilly-sur-Siene, France, 1994, doi:http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA280272.

[41] Meher-Homji, C. B. and Gabriles, G., "Gas Turbines Blades Failures: Causes, Avoidance and Troubleshooting," in "27th Turbomachinery Symposium," Houston, Texas, 1998, pp. 129–180.

[42] Michael Owen, J., "Theoretical Modelling of Hot Gas Ingestion Through Turbine Rim Seals," *Propulsion and Power Research*, Vol. 1, No. 1, 2012, pp. 1–11, doi:10.1016/j.jppr.2012.10.002.

[43] Wilcox, D. C., *Turbulence Modelling for CFD*, DCW Industries, Palm, 1993.